K-MEANS CLUSTERING IN ECB FRAMEWORK

[1]Subhankar Guha, [2]Surendra Nath Bhagat
[1]Asst Professor, [2]Asst. Professor
[1],[2]Department of Computer Science and engineering,
[1]Templecity Institute of Technology and Engineering, Khordha, India
[1]Balasore College of Engineering and Technology, Balasore, India

Abstract:
Nowadays cloud computing technology is becoming a major tool to manage huge data in an efficient manner. Huge amount of data is generated on daily basis and for that reason the data mining algorithm is a required for retrieving essential data as per the need from huge data. As because of huge increase in number of services in the clouds and as well as cloud, so the momentum and importance of cloud computing technology has gained in the field of research and implementation. For efficient data mining performance it is vital to combine distributed data before implement mining algorithm in it. In the following research we have presented in for implementing k means clustering algorithm for the purpose of service discovery from enterprise cloud bus architecture ECB.

I.INTRODUCTION

Cloud have become as a basic infrastructure of computing for the purpose of fast delivery of different resources as utility. Data mining has become popular due to increase of cloud computing usage. Data mining is a process of searching and analysing of different patterns and relations present among the different fields in huge data and retrieving the necessary information which is required for the knowledge base. There are different data mining techniques out of which k means algorithm is one of the popular algorithm for clustering.

There are issues related to performance as there is a sharp increment in the count of clouds and also the services rendered by the cloud. To handle these issues concept of enterprise cloud bus framework has been proposed which can be taken as an abstraction layer of SaaS. In this worka framework is presented for discovery of servicewhere services aregrouped on the basis of k means algorithm and can be complete through the similarity matrix. The required clusters are then transformed to the HUDDI from the cloud enterprise service bus CESB. Through this worknewappropriate data is going to be obtainable in the HUDDI hierarchial Universal Description Discovery and Imtegration and it will be easier to map those data by the scheduler agent.

II.REVIEW OF RELATED WORKS

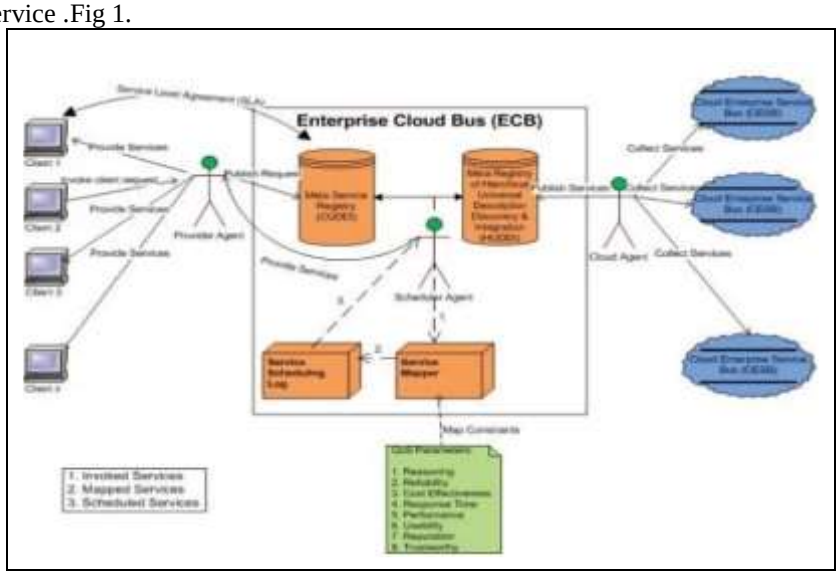
A different Web services can be grouped to form software as a service

Mini present a work on a.SaaS for software system development. There is a demand in service Discovery due to too much increase in software in the cloud . The authors in [1] projected an abstraction layer PCB which works on the basis of multi agent technology and is used in uml 2. In 2 there is a proposal how service can be retrieved dynamically from ECB. In general there are different type approaches are available to give a support in service discovery [3][4,] and in dynamic way in [5][6][7]. In [8] author presenta system whichbring out and find out service on the basis of location and the nomenclature of it is context aware ESB. At this time there is really a need of a new approach where the work will be to monitor and manage services insects depending upon their similarity.

In this research we are try to present a framework which will work on the ECB model.

III.A BRIEF ON ECB

CESB and ECB are thoroughly discussed in [1] and [2]. In SaaS CESB and ECB act as aahigh level of abstraction. The end users of cloud computing environment is our service consumers, the provider agent is use registry CUDDI to publish the client request. A cloud agent is provided by cloud service provider for each cloud, who actually collect the various services from the different CESBs and are published in the HUDDI. There is a deployment of scheduling agent to invoke or discover different services from meta service registry of the Enterprise Cloud Bus (ECB) and then it is published in the mapper and depending on the quality of service Qos parameters the mapper mapping the service .Fig 1.



IV.PROJECTED WORK ON WEB SERVICE DISCOVERY THROUGH CLUSTERING USING KMEANS

Here a framework has been proposed by the help of which most relevant data are clustered. First the user sends a request to the cloud agent about the service according to his requirement.On other hand agent passes the received request to the ECB framework. If the service is available in the HUDDI then it is provided if not available then the request of the user is transfered to the service Discovery module ,this module finds a set of same or similar services and place that cluster to the HUDDI of the ECB framework.

The different servicewhich are available in different CESBs are registered in HUDDI. Here in the following work a framework has been proposed for clustering the services with the help of k means algorithm. As csb is an obstruction for all ESBs that are available in the cloud so the clustering methodology is going to be implemented using k-means algorithm in the CESB.

There are four steps in the algorithm:

1. Initialization.

Data set, centroid and number of clusters are defined for each cluster in the first step.

2. Classification.

The distance is calculated from the centroid for each data point. The data points which are having the minimum distance from the centriod are contained in a cluster.

3. Centroid Recalculation.

The centriods of the clusters that are calculated earlier are recalculated

4. Convergence Condition

A differentconditions of convergence are listed below:

i)It should stop once predefined number of iterations is completed.

ii) Between the cluster if no of data points are exchange then stop.

iii) After achieving the threshold value it should stop.

5. Go back to step 2 if all the above mentioned conditions are not fulfilled and until unless the above mentioned conditions are not satisfied repeat the complete process again.

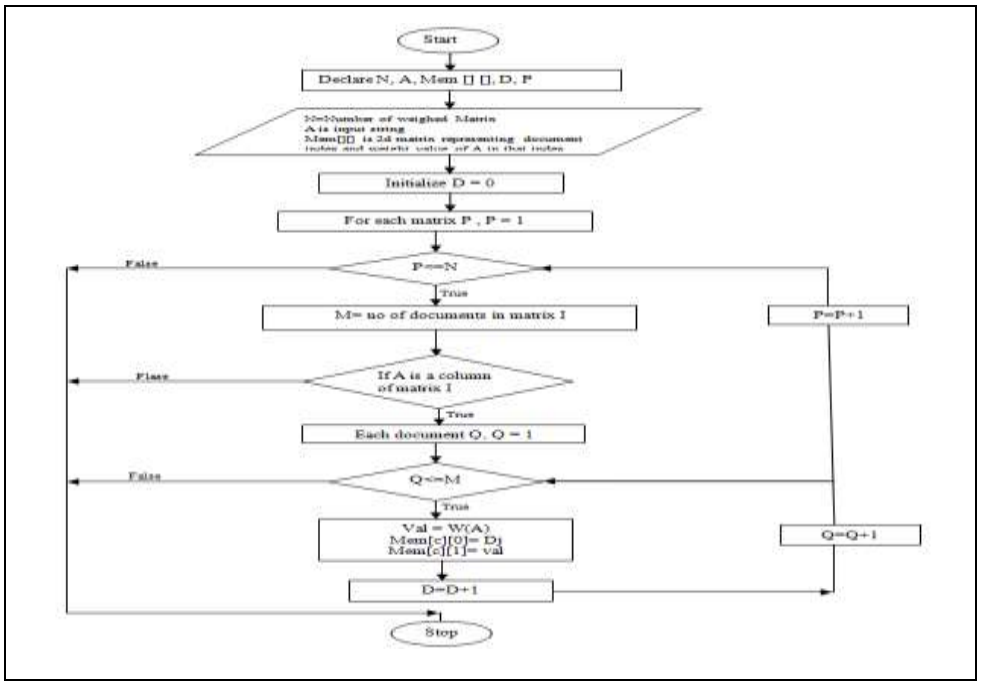
For the information retrivalwors of the data set are required. So tokens are identified which is also known as keywords.Tokenization is the process of fragmenting the sentences into individual tokens, then the weight is calculated using the frequency of the word and by using term frequency-inverse document frequency (tf-idf). TF(t) = (Number of times term t appears in a document) / (Total number of terms in the document). IDF(t) = log (Total number of documents / Number of documents with term t in it). Thus, the Tf-idf weight (Wi)is the product of TF(t) * IDF(t). After calculating Wi a matrix will beformed.

	tw1	tw2	tw3	tw4	tw5
F1	0.5	0.25	0.54	0.125	0.0
F2	0.36	0.35	0.52	0.46	0.35
F3	0.32	0.42	0.52	0.36	0.31
F4	0.25	0.51	0.35	0.25	0.0

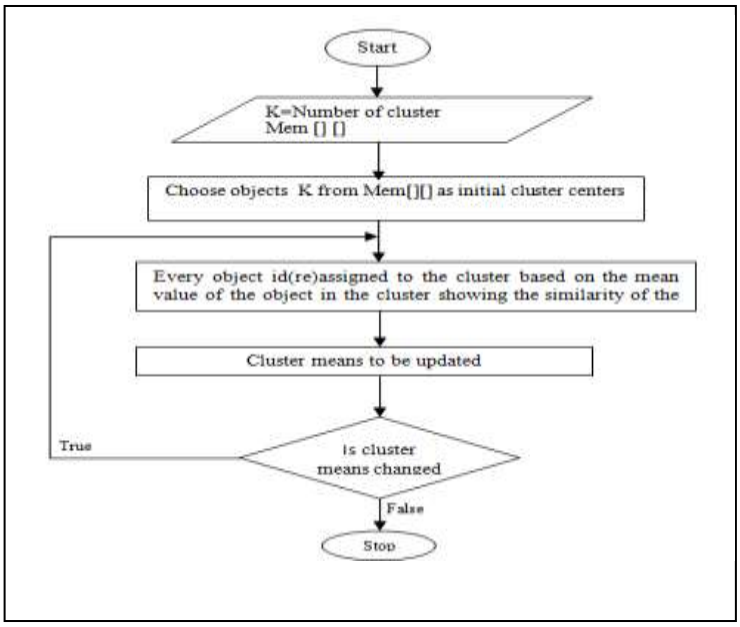
Proposed Floechart for Web Service Discovery in the course of Clustering using K-Means

Part A

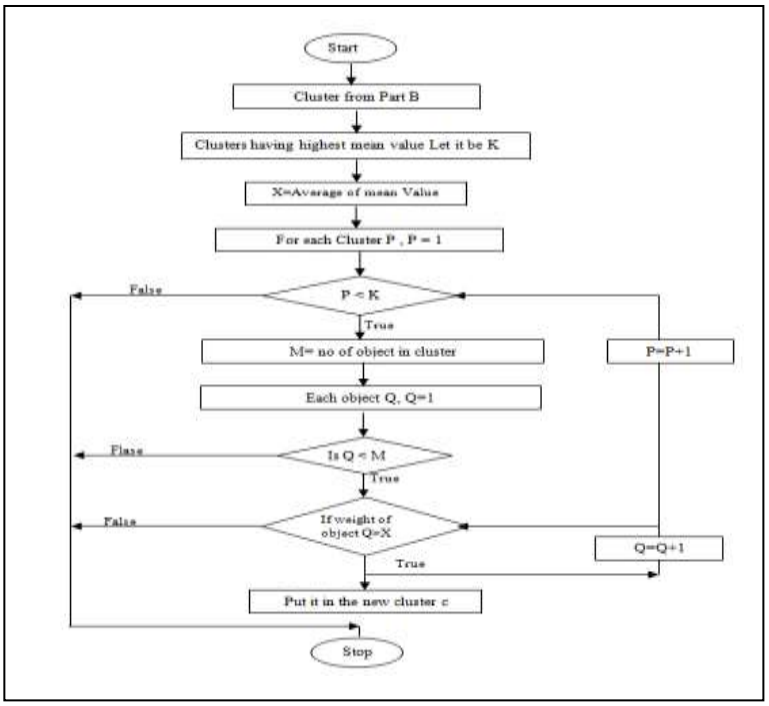
.



Part B



Part C



V.CONCLUSION

The main aim of this work is to decrease the total amount of data that is flowing between the CESB and the HUDDI. Here K- Means algorithm has been used to obtain of relevant information in an efficient manner from a vast collection of different services present in the cloud. This approach can be further developed for improving the time complexity of searching a definite element or service in the cloud.

REFERENCES

[1]. Gitosree Khan, SabnamSengupta, AnirbanSarkar, Narayan CDebnath, “Modeling of Inter-Cloud Architecture using UML2.0: Multi-Agent Abstraction based Approach”, In Proc. Of23rd International Conference on Software Engineering andData Engineering (SEDE), New Orleans, Louisiana, USA, pp149-154, October 13-15, 2014.

[2]. Gitosree Khan, SabnamSengupta, AnirbanSarkar, “WSRM:A Relational Model for Web Service Discovery in EnterpriseCloud Bus (ECB)”, 3rd International Conference on EcofriendlyComputing and Communication Systems (ICECCS), NITK Surathkal, Mangalore, December 18th-21st,2014.

[3]. Li, Chune, Richong Zhang, JinpengHuai, XiaohuiGuo, andHailong Sun. "A Probabilistic Approach for Web ServiceDiscovery." In Proceedings of the 2013 IEEE InternationalConference on Services Computing, pp. 49-56. IEEEComputer Society, 2013.

[4]. Zisman, Andrea, George Spanoudakis, and James Dooley."Aframework for dynamic service discovery." In Proceedingsof the 2008 23rd IEEE/ACM International Conference on
Automated Software Engineering, pp. 158-167.IEEEComputer Society, 2008.

[5]. Vu, Le-Hung, Manfred Hauswirth, and Karl Aberer. "Towardsp2p-based semantic web service discovery with qos support."In Business Process Management Workshops, pp. 18-31.
Springer Berlin Heidelberg, 2006.

[6]. Xu, Jianjun, Qian Zhu, Juanzi Li, Jie Tang, Po Zhang, andKehong Wang. "Semantic Based Web Services Discovery."In Content Computing, pp. 388-393. Springer BerlinHeidelberg, 2004

[7]. Hausmann, Jan Hendrik, Reiko Heckel, and Marc Lohmann."Model-based discovery of Web Services."In Web Services,2004.Proceedings.IEEE International Conference on, pp.
324-331. IEEE, 2004

[8]. Bhattacharya, Swapan, JayeetaChanda, SabnamSengupta, andAnanyaKanjilal. "Dynamic Service Choreography usingContext Aware Enterprise Service Bus." In SEKE, pp. 319-324. 2011